PubH 7401: Elements of Biostatistical Inference I
Homework 3: due Sept. 25

1. Let X be a random variable denoting out-of-pocket health care expenditures in a year for a randomly selected Minnesotan. Define X as follows

$$X = \begin{cases} 1 & \text{if \$0 spent} \\ 2 & \text{if \$0 } < \text{ spent } \leq \text{ \$100} \\ 3 & \text{if \$100 } < \text{ spent } \leq \text{ \$500} \\ 4 & \text{if over \$500 spent} \end{cases}$$

   X is of course discrete. In fact, X is said to be a *categorical* variable with *ordinal* outcomes. The pmf of X is

$$P(X = 1) = 0.1,\ P(X = 2) = 0.2,\ P(X = 3) = 0.45,\ P(X = 4) = 0.25.$$

   (a) What is the probability of not spending anything?

   (b) What is the probability of spending more than \$100?

   (c) What is median of X? What is the mode of X?

2. Diabetes is a growing concern among Inuit, who as a people have gradually shifted to a diet rich in carbohydrates from their traditional diet of seafood. A small town in Labrador (part of a Canadian province) has $n = 243$ Inuit living in it. Let X be the number of Inuit in this town with diabetes. In Labrador, the probability that an individual has diabetes is $p = 0.04$. Assume that individuals have diabetes independently of each other in this town.

   (a) What is the range of X?

   (b) What distribution best describes X?

   (c) What is $P(X \leq 2)$?

   (d) Use the Poisson approximation to estimate $P(X \leq 2)$. How does this compare to (c)?

3. Consider screening a patient for HIV. Let D denote the event that a randomly selected individual from the population has the HIV virus and let S be the event that the HIV screening test comes up positive. The ELISA test has very high sensitivity, about 99.7%, or probabilistically $P(S|D) = 0.997$. The specificity is $P(S^C|D^C) = 0.985$. The prevalence of HIV in the general population is about 0.51%, yielding $P(D) = 0.0051$.

   (a) A randomly selected individual tests positive for HIV; what is the probability that the individual really has HIV? That is, find $P(D|S)$.

   (b) A randomly selected individual tests negative for HIV; what is the probability the individual has HIV? That is, find $P(D|S^C)$.

4. Let X be the survival time in days (after diagnosis) of a randomly selected 60 year-old male diagnosed with small cell lung cancer and treated with a regimen of cisplaten followed by etoposide. The pdf of X is estimated to be $f(x) = 0.00125 \exp(-0.00125x)$ for $x \geq 0$.

   (a) What distribution is this? That is, $X \sim$?

   (b) What is the probability that a randomly selected individual will live less than 30 days?

   (c) What is the probability that a randomly selected individual will live past two years?

   (d) What is the median survival time after diagnosis?

   (e) What survival time do 90% of the population being studied live less than?

5. Chachugi, a particularly good hunter from a Paraguayan Ache' tribe, is going on a three-day armadillo hunting trek. Hunting success contributes to a tribe member's status within the tribe; the number of armadillos killed over time is well-modeled as a Poisson random variable. Say the Chachugi's rate of killing armadillos is $\lambda = 3.5$ armadillos per three days. Let X be the number of armadillos killed by Chachugi over the three day hunting trek.

   (a) What is the probability that Chachugi kills no armadillos over the three days?

   (b) What is the median number of armadillos killed over all treks?

6. The probability of a randomly selected pig being infected with toxoplasmosis is modeled as $X \sim \text{beta}(3, 2)$, for a particular large herd.

   (a) Write down the pdf $f(x)$ of this random variable X. Use the fact that for an integer n, $\Gamma(n) = (n-1)! = (n-1)(n-2)\cdots(3)(2)(1)$.

   (b) Show that the pdf in (a) integrates to one over $R = [0, 1]$.

   (c) Find the cdf $F(x) = P(X \leq x)$.

   (d) Find the mode of X. Hint: the density $f(x)$ has a derivative of zero at the mode.

Also hand in from Chapter 2 in textbook: 34, 40, 46.